Stat 574
11-21-17
(1)

Nonlinear regression

Consider the model $y = \beta e^{\alpha x} + \varepsilon$

$$SSE = \sum_{i=1}^{n} \left(y_i - \beta e^{\alpha x_i}\right)^2$$

$$\frac{\partial SSE}{\partial \beta} = \sum_{i=1}^{n} \cancel{2}\left(y_i - \beta e^{\alpha x_i}\right) e^{\alpha x_i} \overset{set}{=} 0$$

$$\sum_{i=1}^{n} y_i e^{\alpha x_i} = \beta \sum_{i=1}^{n} e^{2\alpha x_i}$$

$$\hat{\beta} = \sum_{i=1}^{n} y_i e^{\hat{\alpha} x_i} \Big/ \sum_{i=1}^{n} e^{2\hat{\alpha} x_i}$$

(2)

$$\frac{\partial SSE}{\partial \alpha} = \sum_{i=1}^{n} \cancel{2}\left(y_i - \beta e^{\alpha x_i}\right)\left(\cancel{-}\beta e^{\alpha x_i} x_i\right) \overset{set}{=} 0$$

$$\beta \sum_{i=1}^{n} y_i x_i e^{\alpha x_i} - \beta^2 \sum_{i=1}^{n} x_i e^{2\alpha x_i} = 0$$

If $\beta \neq 0$, divide both sides by β

$$\sum_{i=1}^{n} y_i x_i e^{\alpha x_i} - \hat{\beta} \sum_{i=1}^{n} x_i e^{2\alpha x_i} = 0$$

& solve for α <u>numerically</u>

③

Ridge regression

Recall $\hat{\beta} = (X'X)^{-1} X'Y$

Consider $\hat{\beta}_R = (X'X + kI)^{-1} X'Y$

k ↑ constant

---

let $MSE(\hat{\beta}) = E[\hat{\beta} - \beta]^2$

"mean squared error"

④

$$MSE(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta}) + E(\hat{\beta}) - \beta]^2$$

$$= E[\hat{\beta} - E(\hat{\beta})]^2 + E[E(\hat{\beta}) - \beta]^2 + 2E[(\hat{\beta} - E(\hat{\beta}))(E(\hat{\beta}) - \beta)]$$

$$= V[\hat{\beta}] + \text{Bias}^2(\hat{\beta}) + 2(E(\hat{\beta}) - \beta)\, E\underbrace{[\hat{\beta} - E(\hat{\beta})]}_{0}$$

So $MSE(\hat{\beta}) = V[\hat{\beta}] + \text{Bias}^2(\hat{\beta})$

5

So, our goal is to find k in ridge regression that will minimize $MSE(\hat{\beta}_R)$

Proposed in 1970 by Hoerl & Kennard

$$E[\hat{\beta}_R] = E[(X'X + kI)^{-1}X'Y]$$
$$= (X'X + kI)^{-1}X' E[Y]$$
$$= (X'X + kI)^{-1}X' E[X\beta + \varepsilon]$$
$$= (X'X + kI)^{-1}X'X\beta$$

6

$$\text{Bias}(\hat{\beta}_R) = E[\hat{\beta}_R] - \beta$$
$$= [(X'X + kI)^{-1}X'X - I]\beta$$
$$= (X'X + kI)^{-1}[X'X - (X'X + kI)]\beta$$
$$= -k(X'X + kI)^{-1}\beta$$

$$V[\hat{\beta}_R] = V[(X'X + kI)^{-1}X'Y]$$
$$= V[(X'X + kI)^{-1}(X'X)\overbrace{(X'X)^{-1}X'Y}^{\hat{\beta}}]$$

$$= (X'X+kI)^{-1} \cancel{X'X} \underbrace{Cov(\hat{\beta})}_{\sigma^2 \cancel{(X'X)^{-1}}} X'X (X'X+kI)^{-1}$$ ⑦

$$= \sigma^2 (X'X+kI)^{-1} X'X (X'X+kI)^{-1}$$

Hoerl & Kennard proposed minimizing the trace of

$$V[\hat{\beta}_R] + [Bias(\hat{\beta}_R)]' Bias(\hat{\beta}_R)$$

They proved (1) There exists a value of k such that $MSE(\hat{\beta}_R) < V[\hat{\beta}]$

(2) As $k \uparrow$, $R^2 \downarrow$

⑧

<u>Logistic regression</u>

$$Y = X\beta + \varepsilon$$

What if Y only takes on the values 0 and 1?

$V[Y] = V[\varepsilon]$, so ε has a Bernoulli distribution, where p is the prob. that $Y=1$

But p depends on X, so $V(\varepsilon)$ is not constant

⑨

What if we try to predict $p = P(Y=1)$ instead of predicting Y?

$$p = X\beta + \varepsilon$$

We may still get predicted values >1 or <0

Try

$$\underbrace{\ln\left(\frac{p}{1-p}\right)}_{\text{logit}(p)} = X\beta + \varepsilon$$

This is called the <u>logistic</u> regression model

i.e., you are predicting the natural log of the odds of success

⑩

$$\text{logit}(p) = w$$

$$\ln\left(\frac{p}{1-p}\right) = w \qquad \frac{p}{1-p} = e^{w}$$

$$p = e^{w} - pe^{w}$$

$$p(1+e^{w}) = e^{w}$$

$$p = \frac{e^{w}}{1+e^{w}}$$

So $\hat{p} = \dfrac{e^{X\hat{\beta}}}{1+e^{X\hat{\beta}}}$

11

How is $\hat{\beta}$ found?

Y has a Bernoulli distribution,

so $P(Y=y) = p^y (1-p)^{1-y}$

The likelihood function is the product of these probabilities

$$L = \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{1-y_i}$$

where $p_i = \dfrac{e^{X_i \beta}}{1+e^{X_i \beta}}$

12

Then, using numerical methods, find the $\hat{\beta}$ which maximizes L.

Stat 564 HW#7

6.1 Perform a thorough influence analysis of the solar thermal energy test data given in Table B.2. Discuss your results.

Note: check my webpage for a link to the Excel file called TableB2.xlsx.

7.6 The carbonation level of a soft drink beverage is affected by the temperature of the product and the filler operating pressure. Twelve observations were obtained and the resulting data are shown below.

Carbonation, y	Temperature, x_1	Pressure, x_2
2.60	31.0	21.0
2.40	31.0	21.0
17.32	31.5	24.0
15.60	31.5	24.0
16.12	31.5	24.0
5.36	30.5	22.0
6.19	31.5	22.0
10.17	30.5	23.0
2.62	31.0	21.5
2.98	30.5	21.5
6.92	31.0	22.5
7.06	30.5	22.5

a. Fit a second-order polynomial.
b. Test for significance of regression.
c. Test for lack of fit and draw conclusions.
d. Does the interaction term contribute significantly to the model?
e. Do the second-order terms contribute significantly to the model?